CHARACTERIZATION OF THE COHESIN COMPLEX IN THE MODEL ORGANISM *NEUROSPORA CRASSA*

by

CHANEY MICHAEL HART

A THESIS

Presented to the Department of Biology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Arts

FALL 2019

An Abstract of the Thesis of

Chaney Hart for the degree of Bachelor of Arts
in the Department of Biology to be taken December, 2019

Title: Characterization of the Cohesin Complex in the Model Organism *Neurospora crassa*

Approved: _____*Eric Selker, Ph.D.*_____
Primary Thesis Advisor

The cohesin complex is a conserved protein complex that plays an important role in multiple aspects of genomic function. Of particular interest is cohesin's demonstrated role in influencing 3D genomic structure. While previous work has identified basic elements of 3D genomic structure in the model organism *Neurospora crassa*, the undermining factors that contribute to these structures are unclear. We hypothesize that the cohesin complex may interact with heterochromatin to shape genomic architecture in *N. crassa*. Features of the cohesin complex such as where it is recruited, its contributions to gene regulation and its presence at topologically associated domains are widely divergent amongst model organisms in which it has been studied, making it important to establish basic features of this complex in *N. crassa*. In this study I took the first steps towards characterizing the cohesin complex in *N. crassa* by showing that cohesin shares features with well characterized yeast species such as enrichment over 3' untranslated regions and intergenic regions of convergent genes across the genome. I also developed a strain of *N. crassa* that has a mutation in cohesin component RAD21 which leads to temperature-sensitive lethality.

My findings and the strains I generated will be useful for further characterization of the cohesin complex in *N. crassa* and for exploration of the role this complex plays in genomic structure and function.

Acknowledgements

I would like to thank the entire Selker lab for helping me in many ways - day to day assistance, guidance on the direction my project and much more. I very grateful to have had the opportunity to be guided by such a knowledgeable, involved and caring primary advisor as professor Eric Selker throughout this project. I also would like to highlight the contributions that Dr. Vincent Bicocca, a postdoctoral fellow in the Selker lab provided. He served as my mentor, trained me in the lab and was instrumental in the intellectual and practical development of my project.

I would like to acknowledge the other members of my thesis committee: my second reader professor Ken-ichi Noma and my Clarks Honors College representative professor Casey Shoop. Professor Noma's support and expertise in my area of research and professor Shoop's assistance in helping to present my work in a way that is impactful to a broad audience were both invaluable. Thank you to Miriam Jordan, the CHC Thesis Coordinator for guiding me through this process in many ways.

Additionally, I would like to thank the Vice President of Undergraduate Research Dr. David Conover and his staff for funding my research over the summer of 2019.

Thank you to my family and friends.

Table of Contents

List of Figures

List of Tables

Introduction

Beyond the DNA sequence

The field of molecular genetics is concerned with structure, regulation and activity of genetic material (material contributing to the transmission of inherited characteristics, collectively called the genome) at the molecular scale. While DNA is the most basic structure that carries genetic information, it does not act alone. It has been known, since the genetic contribution of DNA was elucidated, that other factors are needed to carry out the most basic functions of DNA. For example, the task of expressing genetic information involves transcribing DNA into mRNA and then translating mRNA into proteins that carry out various cellular tasks and contribute to an organism's observable characteristics.

A current frontier in molecular biology is the study of the cellular systems and factors that interact with the genome and carry out genomic functions such as gene regulation, genome defense and repair and the transmission of genetic information from one generation to the next.

Genome structure and function

A common theme in biological systems such as the genome is that structure and function are interrelated. The structure of the genome is important because it is like the gateway to DNA. Processes such as gene expression, DNA repair and DNA replication have a physical tangibility. Thus, the structure of the genetic material that is being acted upon in these processes is important.

Like a tree, the genome has structural attributes at multiple levels (Figure 1). A whole tree has a structure that can be summarized as central trunk with many projecting limbs. Zooming in on one of these limbs shows that the structural pattern of the whole tree is repeated here. Smaller branches emanate from the limb. These emanating branches are similarly subdivided into smaller branchlets and this pattern continues until at the tip of a twig lies a single leaf. Like a tree, the genome has multiple nested layers of structure and the whole genome is an emergent property that is shaped by characteristics of each level. While the structure of the genome is difficult to characterize, modern technological advances have helped to reveal more about how DNA is structured at various levels and about how and why this is important.

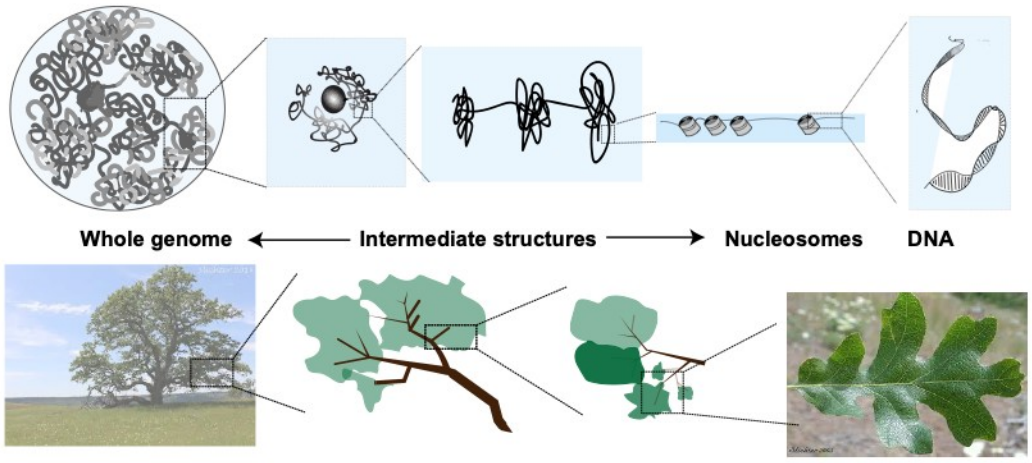


Figure 1: Hierarchical levels of genome structure

The structure of the whole genome is composed of nested sub-levels of structures similar to a tree.

DNA is compacted, condensed and organized within the genome. Without this compaction the genome would not fit inside microscopic cells[1]. At a basal level,

[1]The total amount of DNA within a human cell is 3 million base pairs. Given that the distance between base-pairs is 3.4 angstroms, the size of the human genome stretched out lengthwise would be 1 meter. Most human cells are diploid meaning they contain

strands of DNA are wound around nucleosomes[2] into a fiber which resembles beads on a string. The collective name for DNA wrapped around histones and associated with other nuclear proteins and RNAs is chromatin. Chromatin is condensed and organized into discrete spans that are called chromosomes which can be seen with the aid of a high power microscope.

While the organization of the genome between the scale of the nucleosome and whole chromosomes is an active area of research, it is known that that many levels of structure that are functionally important exist within this range. Within non-dividing cells the chromatin is more relaxed and it is difficult to judge with the eye where one chromosome ends and another one begins. While this may initially suggest that the whole genome lacks an ordered structure, a well-established body of work demonstrates that the genome has an architecture – a structure that is optimized to enable genomic processes and functions to flow with maximum efficiency.

In a very simple genome it is theoretically possible to optimize the efficiency of transcription, replication and other processes by optimally ordering genes and regulatory elements along the 2D span of a chromosome. However, even in a relatively simple single cell eukaryote such as a yeast, the genome is too large and contributes to too many processes for 2D organization to accommodate. The genome, however, is not linear but has three dimensions and this greatly increases the possibilities for efficient compartmentalization of genomic regions according to function. For example, actively

---

two copies of the genome. Thus the total distance would be 2 meters.

[2] A nucleosome is composed of an octamer of histone proteins. Histones are small positively charged proteins that are well conserved throughout the eukaryotic domain. There are five members of the histone protein family: H1, H2A, H2B, H3 and H4. All but H1 are typically found in nucleosomes while H1 is thought to contribute to higher order structure.

transcribed regions could be segregated away from inactive regions and these broad domains could contain smaller domains that form between regions that tend to be transcribed at the same time or are controlled by common gene regulation pathways. In addition, these domains may be dynamic, as the same gene might be grouped differently in different cell types, in different developmental states or in response to external stimuli.

Recently developed techniques provide ways to test hypotheses about how the 3D genome is organized.[1,2] Chromosomal Conformation Capture (3C) is a method that assays 3-D genomic structure. In this technique genomic regions that are in spatial proximity within the nucleus are crosslinked and then ligated to each other. The frequency with which two loci interact (ligated to each other) can be found. And this frequency is correlated to distance in space[1]. Hi-C builds on the 3C method and takes advantage of high-throughput next generation sequencing technology to assay the interaction frequency of every genomic locus (at a given resolution) in reference to every other locus in the genome.[3] From this, an interaction matrix can be constructed which provides an estimation of how the genome is organized in 3D space.

Data from Hi-C experiments have shed light on common features of 3D genomic architecture that are summarized in Figure 2.[2,4–6,7,8,9]

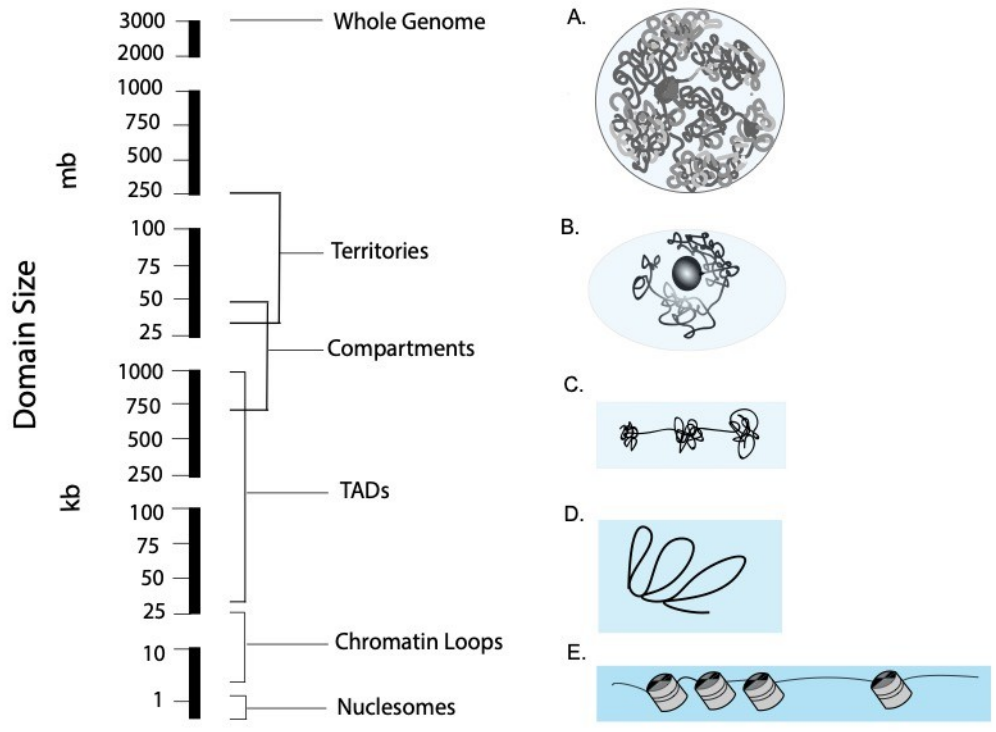


Figure 2: Features of 3D genome structure

Technological improvements such as Hi-C have revealed features of genomic structure at different scale levels. The whole genome (A) is segregated into chromosome territories. Within and occaisonally between territories are nested substructures including compartments (B), topologically associating domains, or TADs (C) and chromatin loops (D). At a basal level chromatin is made up of DNA wound around nucleosomes (E). This Schematic shows an approximation of the structure of these features and their size based on Hi-C experiments done in human cells.

Each chromosome is segregated into distinct territories. Thus globally, intrachromosomal contacts are more frequent than interchromosomal contacts.[2] It can be inferred from a principal component analysis of interaction frequencies that two primary compartments exist (called the A and B compartments). Every locus in the genome belongs to one of these compartments and tends to associate more with loci that are members of the same compartment rather than members of the other compartment.

Compartments are strongly correlated with alternate forms of chromatin called euchromatin (A compartment) and heterochromatin (B compartment). Euchromatin is a relaxed form of chromatin that contains the actively transcribed parts of the genome. Heterochromatin is a form of chromatin that is thought to be a tightly condensed and forms over gene poor regions, especially the centromeres and telomeres[3]. It plays roles in major genomic functions such as gene silencing and genome defense. It is associated with modifications at the sub-nucleosomal and nucleosomal scale such as DNA methylation, histone 3 lysine 9 methylation (H3K9me) and/or histone 3 lysine 27 methylation (H3K27me)[4].

Hi-C data have also highlighted the existence of topologically associated domains (TADs) and chromosome loops. TADs are identified as collections of sequences that associate with themselves more than with any other sequence outside of that collection.[10] TADS are distinct from compartments.[11] While a compartment is made up of sequences that associate preferably with other members of the same compartment, a TAD is made up of sequences that associate with each other more than with any other regions in the genome. TADs are thought to represent compact aggregations of chromatin based on 3D modelling and are sometimes referred to as "globules."[4] Chromosome loops are inferred from point interactions between loci that have a much

[3] Each chromosome has specialized segments called the centromere and the telomeres. The centromere is often found near the center of the chromosome. It is where replicate chromosomes are joined together during mitosis and meiosis. The telomeres are found at the tips of each chromosome. They are important for protecting chromosomes in various ways.

[4] DNA can be chemically modified by the addition of a methyl group onto cytosine, one of the four constituent bases in DNA. Additionally, histones can be chemically modified in a variety of ways. Both DNA and histone modifications are correlated with genomic processes such as transcription, gene regulation and protection of the genome.

higher interaction frequency than expected given their chromosomal distance.[6] Often chromosome loops are contained within or form the boundaries of TADs.[6]

While the investigation of the functional importance of these structural domains is still in early stages, 3D genomic structure has been shown to play an important role in the development and maintenance of cellular regulatory networks. By forming globule-like domains, TADs partition the genome into distinct chromatin neighborhoods. Within these neighborhoods enhancers and other regulatory sequences are spatially confined so that they can only interact with genes also in that TAD. Disruptions of TAD boundaries has been shown to be the root of some identified developmental defects and to contribute broadly to oncogenesis in humans and mammal models.[12–14] It is plausible that these pioneering studies are the tip of the iceberg and genomic structure will be found to have a direct and broad impact on development and gene regulation.

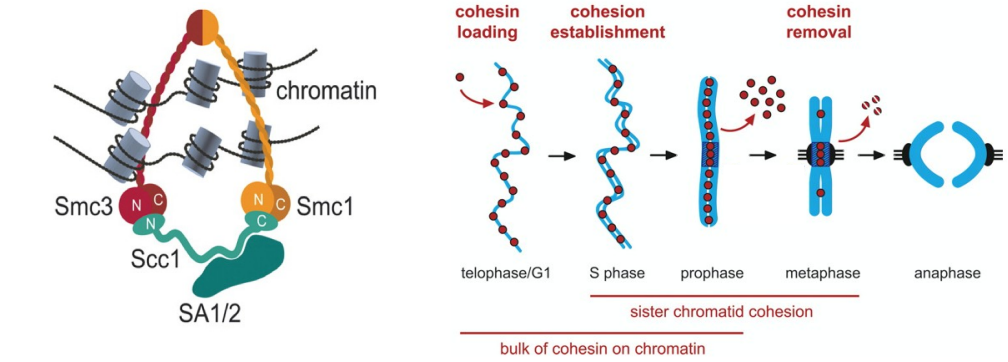
The cohesin complex and genomic structure.

Given that the organization and structure of the genome is not random, another important question is what factors influence this structure. The cohesin complex, amongst other complexes which contain SMC (structural maintenance of chromosome) family proteins, has a demonstrated role in influencing genome organization in various ways. The cohesin complex is made up of four core proteins: SMC1, SMC3, RAD21 and SA1/2.[(5)] Cohesin was first characterized as an essential factor for sister chromatid cohesion[(6)] during DNA replication and cell division.[15,16] Cohesin is able to impact the

[5] These are the names of the cohesin complex components in humans

[6] Before cell division (mitosis) each chromosome is replicated. During mitosis, the replicates (called sister chromatids) are held together and oriented in the center of the nucleus. An important step in ensuring that each cell has an entire copy of the genome is the segregation of sister chromatids – one replicate of each chromosome is sent to

physical structure of the genome due to its ring-like structure, shown in Figure 3, which can topologically entrap a strand or several strands of chromatin fiber.[17,18]



Jan-Michaels Peters *et al*. 2008

Figure 3: The structure and mitotic role of the cohesin complex

A. Structure of the cohesin complex. Protein names here are from humans.
B. Cohesin is essential for its role in establishing sister chromatid cohesion during cell division. The model presented here represents cohesin dynamics in humans.

Beyond its role in cell division, cohesin impacts genome structure and function in interphase cells. Mutations in the cohesin complex and in NIPBL, a protein that is required for the loading of cohesin, were identified as the causative mutations in Cornelia de Lange syndrome, which is characterized by developmental anomalies in limbs, short stature and intellectual development defects.[19,20,21] Cohesin's role in chromosome segregation is not responsible for the changes in gene regulation in Cornelia de Lange syndrome.[22–25] Additionally, non-mitotic functions of the cohesin

each daughter cell. Cohesin is essential because it is coheres sister chromatids, it helps to orient these sister chromatids in the center of the nucleus properly and its cleavage is necessary for sister chromatid segregation.

complex have been shown to be essential for the proper development of neurons and other cell types in the fruit fly, *Drosophila melanogaster*.[26–28]

The chromosomal binding "addresses" of the cohesin complex correlate with the boundary points of TADs and cohesin has been shown to be necessary and for TAD formation and/or maintenance in many organisms.[4,29–32] Cohesin also mediates the formation of chromosome loops that bring together factors such as promoters and enhancers to promote gene expression.[6,33–36] Cohesin is thought to establish TAD domains and loops by extruding chromatin through its ring.[30,37,38]

Additionally, cohesin has been implicated in interacting with heterochromatin to determine the genomic architecture of the yeast *Schizosaccharomyces pombe*. Importantly in Hi-C analysis of Δclr4 (the gene that encodes *S. pomb*e's H3K9 methyltransferase) global loss of contact between heterochromatin regions was reported alongside of decreased recruitment of cohesin to centromeres.[4]

Despite evidence that the cohesin complex contributes to important aspects of genome structure many questions remain regarding the cohesin complex. The mechanism underlying cohesin's role in TAD and chromatin loop structures, the biological ramifications of these cohesin-associated structures and the relationship between these structures with other features such as histone modifications and alternate chromatin states remains unclear.

Genome function in *Neurospora crassa*

*Neurospora crass*a is a well-studied and important model organism that has been exploited for research in a wide range of topics including but not limited to basic

genetics, biological clocks and epigenetic processes.[39] *N. crassa* is a multicellular filamentous fungus in the ascomycota phylum. In its natural environment, *N. crassa* can be found in most parts of the world.[40] It is commonly seen germinating in burned wood after a forest fire as *N. crassa* ascospores require heat to germinate.

While it is distantly related to humans and other animals, at the cellular level, there is a great deal in common between *N. crassa* and other eukaryotic organisms (such as humans). Common features of all eukaryotes include the storage and organization of DNA in chromatin within the nucleus, the presence of membrane-bound organelles such as mitochondria and the ability for many cells to form multi-cellular organisms. Basic processes such as transcription, translation, gene regulation and cell division are common to all eukaryotes.

N. crassa is an especially powerful system for the study of many aspects of genomic function including epigenetics and gene regulation that are of interest in human biology. Many features that characterize complex mammalian chromosomes such as large heterochromatin domains, DNA methylation and histone modifications (notably H3K9me3 and H3K27me) are also present in *N. crasssa*. In addition, *N. crassa* is well suited to the study of these chromosomal features and their implications because it has a relatively small genome[7], it replicates quickly, it is haploid for most of its life[8] and methods for genetic and proteomic manipulation of *N. crassa* are well established.[41,42] In the Selker lab, we take advantage of *N. crassa* as a model organism to

[7] *Neurospora's* genome size is ~40 Mb. In comparison the human genome is ~3000 Mb.
[8] This means that there is only one copy of the genome present in the cell for most of *N. crassa's* life cycle. This is beneficial from an experimental standpoint because it means that most genes are essentially monoallelic. If a gene is mutated or knocked out of the genome then there is no other copy of this gene that will compensate for it.

investigate epigenetic processes, gene silencing, chromatin structure/function and other topics relating to genomic function.

The 3D genomic structure of *N. crassa* is demonstrably highly organized.[43,44] Strong genome wide interactions among heterochromatin regions is the most prominent feature of 3D genome structure in *N. crassa.*[43] However the deletion of two proteins that are essential for the propagation of the H3K9me3 mark that underlies constitutive heterochromatin: DIM-5 (*N. crassa*'s H3K9 methyltransferase) and HP1 (heterochromatin protein 1) results in only mild reduction in heterochromatin interactions. This suggests that heterochromatin in and of itself is not sufficient for these interactions and that undetermined factors play an important role. Given that cohesin influences genomic structure in other organisms we hypothesized that cohesin may be an important factor for the genomic structure of *N. crassa* as well.

In this study I take the first steps towards characterizing the cohesin complex in *N. crassa*. Because aspects of cohesin vary greatly in the model organisms in which it has been studied, it is necessary to characterize basic features of the complex in *N. crassa.* My findings will provide an essential backdrop for the investigation of cohesin's impact on 3D genomic structure in this organism. I aim to show where cohesin locates in the genome and to determine if patterns of chromosome association are shared with other model organisms. I also aim develop a strain of *N. crassa* that has a loss-of-function mutation in cohesin that can be used for future studies.

Divergent characteristics of the cohesin complex

The basic structure of each of the components of the cohesin complex are evolutionarily ancient and relatively conserved across the whole eukaryotic domain. In

addition, cohesin's role in faithful chromosome segregation is thought to be conserved and essential for all eukaryotic organisms. Cohesin's chromosomal association has been assayed via a method call chromatin immunoprecipitation (ChIP) in *S. cerevisiae* (budding yeast), *S. pombe* (fission yeast), *D. melanogaster* (fruit fly), *M. musculus* (mouse) and *H. sapiens* (human). No universal common mechanism, rule or pattern satisfactorily explains or underlays cohesin's chromosomal binding sites across the genome in any of these organisms, but some shared features exist. Within the previously mentioned organisms cohesin has a broad enrichment profile - it localizes to both heterochromatin and euchromatin regions and it's initial loading is dependent on the activity of the cohesin loading complex. Beyond these commonalities there is considerable variation in other aspects of cohesin's chromosomal association pattern which are summarized in Table 1.

Table 1: Cohesin chromosome association in various model organisms

Model	Correlates with					
	Convergent genes	3' end of genes	Active genes	Promoter regions	Cohesin loading complex	CTCF binding sites
S. cerevisiae	Yes, strongly	Yes	No	No	No	NA
S. pombe	Yes	Yes	Moderately	No	Moderately	NA
D. melanogaster	No	No	Yes, strongly	Yes	Yes	No
H. sapiens	No	No	Yes	No	No	Yes
N. crassa	?	?	?	?	?	NA

In *S. cerevisiae*, cohesin is largely excluded from gene bodies and localized at intergenic regions between the transcription termination sites of genes that are in a convergent orientation (are transcribed on opposite strands of DNA and in opposite directions relative to each other).[45–47] Similar patterns but to a lesser degree are observed in *S. pombe*. In both yeast species cohesin does not share binding sites with the cohesin loading complex except in cases where transcription is abolished suggesting that the process of transcription displaces cohesin from its original loading site and deposits it in intergenic regions.[46,48] In mammalian cells, cohesin co-localizes with an insulator protein, CTCF, that is implicated in multiple gene regulation events but is not significantly linked with gene bodies, gene promoters or intergenic sites.[49–51] Additionally, while cohesin is not enriched at intergenic regions between convergent genes, it is over-represented at sites where two CTCF proteins, which have distinct front sides and tend to form dimers, are in a convergent orientation.[6] In *Drosophila melanogaster* the pattern is again different. Cohesin does not associate with CTCF (*D. melanogaster* has CTCF while *N. crassa* and both yeast species do not). It also co-

localizes with the cohesin-loading complex and both cohesin and its loader have a strong preference for active genes, especially with the promoters[9] of active genes.[52]

Why cohesin has such divergent characteristics in different organisms is unclear. It is possible that cohesin has evolved from its evolutionary ancient role in chromosome segregation to serve other purposes and its chromosome association patterns vary because of this. The size difference between unicellular yeast and both *D. melanogaster* and mammals is vast. Uses for cohesin that were not necessary in smaller genomes may have been selected for as genomes became larger. Filling in the picture of cohesin in *N. crassa* will add additional perspective to the broad picture of cohesin's function. *N. crassa*'s genome, at 40 million base-pairs is substantially larger than that of either yeast (~12 Mb for *S. cerevisiae* and ~14 Mb for *S. pombe)* while it is smaller than *D. melanogaster's* 140 Mb genome.

Rad21-K1

Due to of cohesin's essential role in mitosis it is not possible to generate cohesin knockout mutants. However, in fission yeast, a screen was used to identify a temperature-sensitive loss-of-function mutant defective in cohesin sub-component RAD21. The wild type copy of rad21 was replaced by a randomly mutagenized copy linked to a selectable marker. One recombinant, called rad21-K1, was viable at the permissive temperature (26˚C) but not at 36˚C or above.[53] Above the restrictive temperature, this mutant strain exhibited chromosome segregation defects that were manifested in various abnormal mitotic morphologies such as lagging chromosomes,

[9] The promoter is a region of gene located upstream of the transcription start site. Promoters work with other regulatory sequences, proteins and RNAs to initiate transcription.

unequally segregating chromosomes and cells that had begun separating before nuclear division[(10)]. Future studies established that sister chromatid cohesion is impaired in this mutant background.[54,55]

While the phenotype of rad21-K1 is well-established, details of the mechanism for this phenotype – such as whether cohesin fails to load, whether it loads but then disassociates or is degraded or whether it is fully established but is unable to fulfil its normal function are not clear. In asynchronous cells, viability decreases immediately after shift to the restrictive temperature but cells continue to grow for at least two cell cycle rounds.[53] When cells are released synchronously from G1 to the restrictive temperature viability decreases after 2 hours.[54] This suggests that cohesin becomes essential during S phase[11] (which takes place from 2 to 4 hr after the shift).[54] Rad21-K1 is hypo-phosphorylated.[54] Rad21's cell-cycle dependent phosphorylation by CUT-1 is necessary for separase mediated cleavage at anaphase.[56] Hypo-phosphorylation could indicate that cohesin is not present at the cell cycle stage in which phosphorylation normally occurs. Cohesin is required for kinetochore-microtubule[(12)] interaction in yeast

---

[10] The cell cycle is an ordered sequence of events that occurs in a cell in preparation for cell division. The cell cycle has four stages (G1, S, G2, and M). G1 is the stage of the cell cycle in which the cell grows to become large or healthy enough to divide. S phase is the cell cycle stage in which the DNA within the nucleus is replicated. In G2 the cell makes additional preparations for cell division. The M stage, or mitosis has four steps. Briefly, in prophase condensed sister chromatids are assembled. In metaphase sister chromatids align in the center of the nucleus. They segregate to either side of the nucleus in anaphase after cleavage of the cohesin-mediated link that holds them together. In telophase and cytokinesis the nucleus and surrounding cell of the original cell is divided into two daughter cells, each with a full set of the genome.

[11]

[12] The kinetochore is a protein complex that helps to position sister chromatids during cell division. It does so by associating with microtubules which are part of a network called the cytoskeleton which helps to mediate structure and motility within the cell.

and vertebrates and this interaction is abolished in the rad21-K1 background.[55,57–59] This activates the spindle assembly checkpoint which puts a block on division. Decreased vitality may be a result of cells being blocked before division. In cases where the checkpoint is ignored, chromosome mis-segregation ensues.

While multiple mutation sites exist within the *S. pombe* rad21-K1 strain that was originally developed, the mutation responsible for this strain's phenotype was identified from suppressor screening as a change from the 67th amino acid isoleucine to phenylalanine.[60] This amino acid lies within the N-terminal domain of RAD21 that interacts with the coiled-coil domain of PSM3(SMC3). The existence of many suppressors for this mutation in the corresponding interacting region of SMC3 suggests that the interaction between RAD21 and SMC3 is de-stabilized in the rad21-K1 background. However, whether these proteins are fully or partially separated is not known.

Research Techniques

Epitope tagging.

The first step I took in characterizing the cohesin complex was the addition of a small tag called a FLAG peptide to the C-terminal end of two components of the cohesin complex NCU01247, the *N. crassa* homolog of PSC3 (SA1/2 in humans) and NCU03291, the *N. crassa* homolog of RAD21 (I will refer to these proteins by their name in *S. pombe:* PSC3 and RAD21).[61] This was done by creating a linear segment of DNA that contains the sequence for these genes in addition to sequences for three FLAG peptides, a flexible chain of 10 glycines and a gene (nat-1) that confers resistance to the antibiotic nourseothricin.[42] This sequence was inserted into *N. crassa.* In a subset of the culture into which I inserted the construct, the altered psc3 and rad21 segments were swapped with the wild type copy in a process called homologous recombination. Individuals that took up my altered copy could be selected for by plating them on medium that contains nourseothricin. Because wild type *N. crassa* has no resistance to this drug, only individuals that took up the construct could survive.

ChIP

The addition of the FLAG peptides aided in a downstream analysis called chromatin immunoprecipitation (ChIP). ChIP is used to determine where features interact with DNA sequences genome wide. In my case I used ChIP to assay where my proteins of interest preferentially interact within the genome. This information is useful in determining a protein's function. In this technique formaldehyde is used to cross-link a protein of interest throughout the nucleus to the genomic regions that they are closest

to in 3D space. Interactions between my protein of interest and DNA are isolated by incubating samples with an antibody that recognizes and binds the FLAG peptide. Following stringent washes and reverse cross-linking I was left with the genomic sequences that correspond to where my protein was present. One of two things are done with these sequences. In ChIP qPCR analysis, I probe whether a predetermined sequence is enriched within the sequences pulled down by the antibody. High enrichment of this sequence would indicate that my protein was associated at that site. In whole genome sequencing (ChIP-seq) I prepare all of the DNA from the antibody pull-down into a library. The sequence of all of the DNA in this library is determined using next-generation sequencing technology. From this data I can cross-reference the sequenced regions in the library to the known and properly ordered whole genome sequence of *N. crassa* to see which genomic regions are enriched with my protein. In both ChIP qPCR and ChIP-seq there is a considerable amount of background noise. In order to have confidence that my results indicate a real enrichment of my protein, the expected signal must be significantly higher than the background. In both experiments negative controls of various sorts are helpful in distinguishing signal from noise.

Site-directed mutagenesis.

Site-directed mutagenesis is a method used to introduce changes in the coding sequence of a gene that result in a change in that gene's protein product. To accomplish this, primers are made in which the native sequence is altered to bring about the desired mutation. These primers are designed so that they can be stitched into the wild type copy of the gene without disrupting any additional sequences. The altered gene sequence is attached to a 10xGly3xFLAG::nat-1 cassette and introduced into *N. crassa* in the same way as in FLAG-tagging a non-mutated protein.

Results

Cohesin localizes preferentially to 3' end of genes and to convergent genes in *N. crassa*.

I generated a strain in which three FLAG peptides were tethered to the C-terminal end of PSC3 that is expressed alongside nat-1. Genomic incorporation and protein expression of the psc3-FLAG construct was verified by PCR amplification over the incorporated construct and western blot respectively. I performed ChIP-seq on psc3-3xFLAG using Sigma ANTI FLAG affinity gel (A2220). A wild type strain without FLAG was used for a control (mock ChIP). Robust peaks where signal is several times greater than noise were not observed. In contrast to cohesin ChIP results in other model organisms, cohesin enrichment was not higher at centromeres than along euchromatin arms. Cohesin has been reported to be a difficult protein complex to capture via ChIP in other organisms because it localizes very broadly, and unlike other proteins that tightly chemically interact with DNA, cohesin's topological interaction with DNA is relatively loose.[62] Thus the broad distribution and low signal to noise ratio of my ChIP data was not unexpected. While the quality of my data made it difficult to define discrete peaks of cohesin enrichment, through genome-wide analysis of patterns in psc3-FLAG and mock FLAG ChIP seq data, I determined specific characteristics of PSC3 association.

I first analyzed whether cohesin binding was associated preferentially with genomic features such as gene bodies or 5' or 3' untranslated regions[13]. Using

[13] The part of the gene that is transcribed into mRNA is referred to as the gene body. Genes also contain sequences that are not either neither transcribed nor translated or are transcribed but not translated. Often these sequences play roles in regulating the gene. I was interested in determining cohesin's binding to regions upstream of the transcription start site (5') and downstream of the transcription end site (3') of genes across the

ComputeMatrix from the open source bioinformatics platform Galaxy, I averaged the number of reads from psc3-FLAG and mock FLAG ChIP-seq over every gene in the *N. crassa* genome.[63] Each gene was scaled to 2000 base pair regions (meaning that small genes were stretched out to meet 2000 bp and large genes were condensed) An additional 2500 bp upstream and downstream of the gene body was also analyzed. Galaxy's ComputeMatrix averaged the values over this span into 50 base pair bins. Using Galaxy's PlotHeatmap and PlotProfile the average number of reads (normalized to reads per kilobase per million) of every gene was visualized in meta-plots. Reads were also mapped to genes separated into quartiles based on their expression level in a wild type background. The first quartile contains the most highly expressed genes, the second quartile contains the 50th to 75th percentile of highly expressed genes, and so on.

Meta-plots showed that cohesin is highly enriched over the 3' end of genes. (Figure 4). This is consistent with cohesin association sites in *S. pombe* and *S. cerevisiae*. In addition, the degree to which this 3' bias was observed was correlated with expression quartile (Figure 4B). The 4th quartile (silenced and lowly expressed genes) had a more uniform enrichment profile and the proportion of enrichment over the 3' end increased in the next quartile. The 1st quartile (most highly transcribed genes) showed this bias the most clearly. This suggests that there may be relationship between cohesin positioning and transcription in *N. crassa* as is demonstrated in yeast. In both yeast species the current model is that cohesin is loaded to the 5' end of genes, often near the promoter region and then RNA polymerase physically pushes cohesin downstream during transcription.[46] Cohesin is thought to preferentially localize at

genome.

convergent loci because here cohesin complexes, pushed from both directions, accumulate. It has also been suggested that cohesin plays an active role at transcription end sites. In particular it is thought to promote transcription termination.[47]

It is possible that uniform signal across the fourth quadrant was at least partially due to incomplete or erroneous annotation of genes. Annotation of genomic features such as transcription start sites and end sites in *N. crassa* is informed by RNA-seq data and thus there is little or poor data for genes that do not express much mRNA.

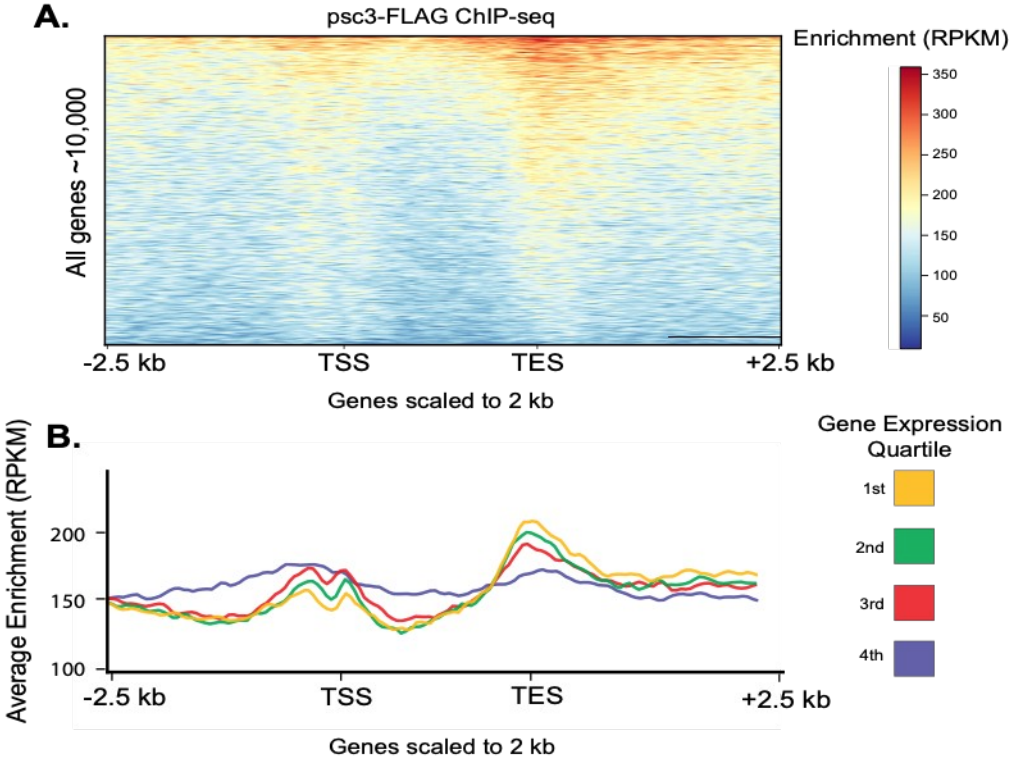


Figure 4: Cohesin enrichment over transcription end sites across the *N. crassa* genome varies with transcription level

A. Heatplot shows psc3-FLAG enrichment over all genes in the *N. crassa* genome. Genes are arrayed in order of average enrichment (highest to lowest) along the Y axis. X axis shows how enrichment aligns to each gene. Color of heatmap corresponds with enrichment strength.

The correlation of cohesin with the 3' end of genes prompted me to investigate if the set of genes that are enriched at their 3' ends tend to be orientated in a particular way. Specifically, I was interested in whether cohesin is preferentially associated with convergent gene orientation as is observed in *S. cerevisiae* and *S. pombe.*

I generated a file which lists every gene pair that is arranged in a convergent orientation as well as files for all divergent gene pairs and for all gene pairs that point in the same direction. Overall there are 2680 convergent gene pairs, 2682 divergent gene pairs and 4361 same direction gene pairs. The deviation from an expected 1:1:2 ratio via the chi^2 statistic is significant ($p<0.001$) suggesting that convergent genes may be preferred by *N. crassa*. Using Galaxy's ComputeMatrix, I generated a matrix that scored cohesin ChIP values over the regions between every pair of genes in the *N. crassa* genome. Cohesin ChIP values reflected the $\log_2$ difference between the number of psc3-FLAG reads and mock FLAG reads averaged in 50 base pair bins. Intergenic regions that had a $\log_2$ ratio of 1.0 (double that of mock FLAG) on average or greater were categorized as being cohesin-bound. The overlap between cohesin-bound gene pairs and convergent, divergent and same direction gene pairs was visualized and quantified using BioVenn.[64] The hypergeometric distribution was used to determine the statistical significance of overlap.

Using this criterium, 50.7% of the intergenic regions between convergent gene pairs were cohesin-bound, which was significantly more than expected. ($p = 2.02 \times 10^{-23}$) (Figure 5). 8.4% of divergent peaks were enriched in cohesin which was significantly lower than expected ($p = 3.23 \times 10^{-22}$). The enrichment of cohesin over genes

transcribed in the same direction in the negative direction was not significantly different than expected. This suggests that as in *S. cerevisiae* and *S. pombe*, cohesin preferentially associates with convergent genes in *N. crassa*. Based on these findings it is plausible that cohesin positioning in *N. crassa* fits under the transcription-based model that has been demonstrated in yeast.

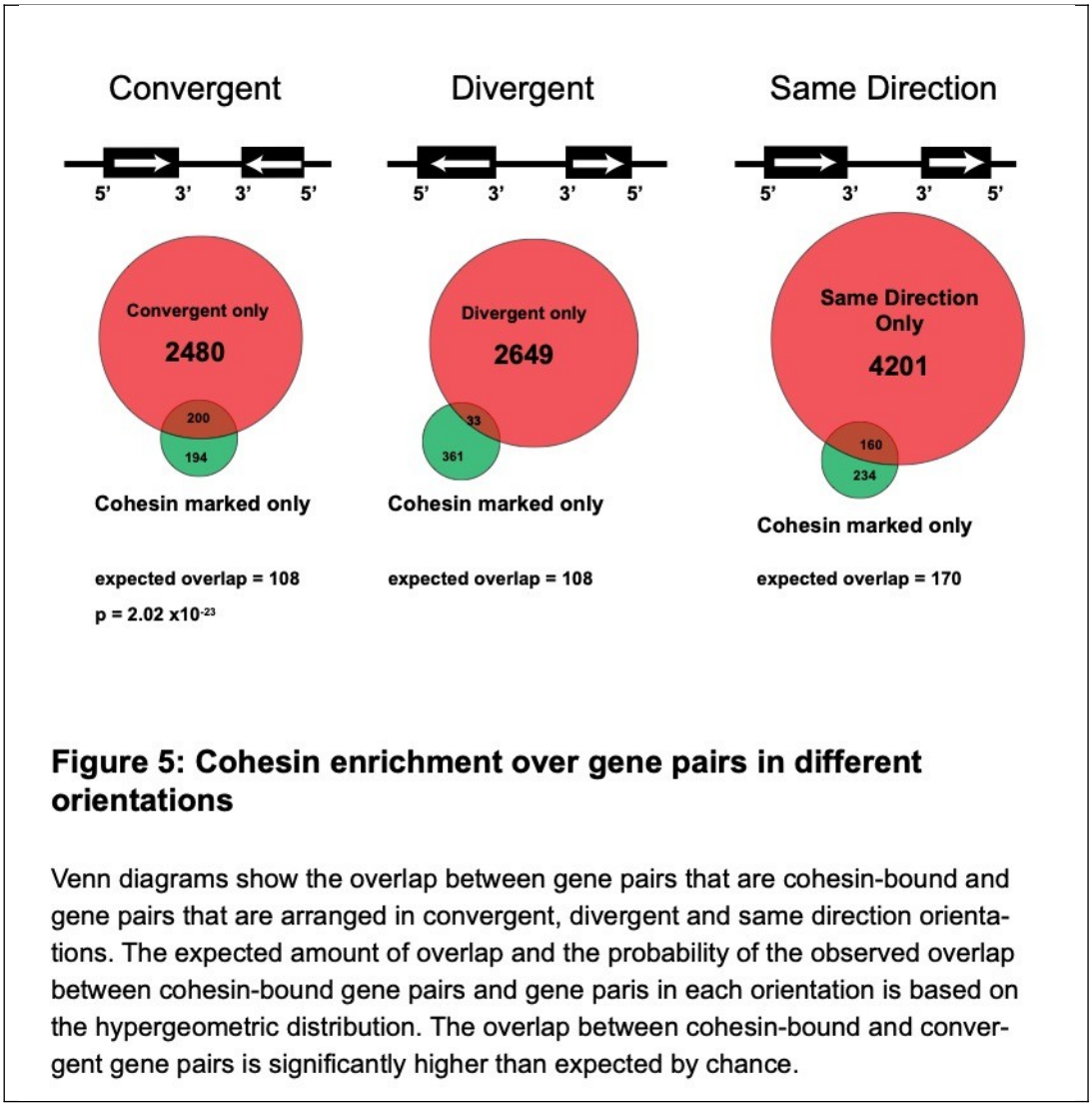


Figure 5: Cohesin enrichment over gene pairs in different orientations

Venn diagrams show the overlap between gene pairs that are cohesin-bound and gene pairs that are arranged in convergent, divergent and same direction orientations. The expected amount of overlap and the probability of the observed overlap between cohesin-bound gene pairs and gene paris in each orientation is based on the hypergeometric distribution. The overlap between cohesin-bound and convergent gene pairs is significantly higher than expected by chance.

Critical domains of RAD21 are conserved in *N. crassa*.

Several regions are well conserved between RAD21 in *N. crassa* and RAD21 in *S. pombe*, *S. cerevisiae*, *D. melanogaster* and *H. sapiens* as shown from NCBI's algorithm for constraint-based-alignment (COBALT). These regions correspond to interaction sites with SMC3 (red, Figure 6), (PSC3) (green) and SMC1 (blue).[65] The causative mutation of the rad21-K1 strain (I67F) lies within the N-terminal domain of RAD21 that is integral for interaction with SMC3 and is well conserved amongst these species and in *N. crassa.*

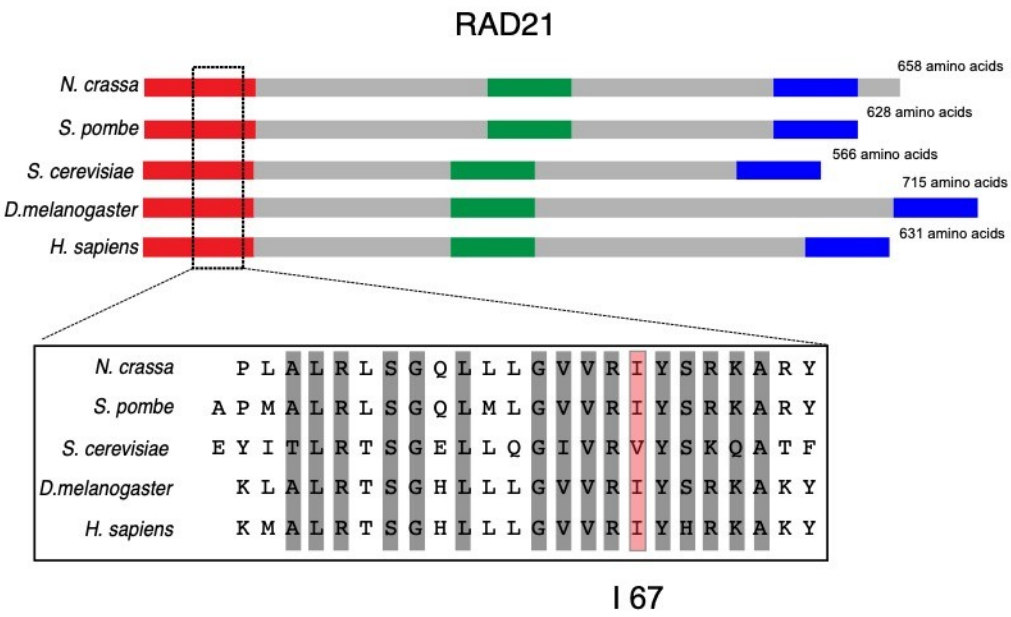


Figure 6: RAD21 homologs

Schematic shows the simplified COBALT alignment between RAD21 in *Neurospora crassa* and four other widely studied organisms. Isoleucine 67 itself is conserved and it lies is within the well conserved N-terminal domain of rad-21 that interacts with cohesin component SMC3. Other conserved regions include the interaction site with PSC3 (green) and SMC1 (blue).

Temperature sensitive lethal RAD21 mutation isolated in *N. crassa*.

I designed primers to incorporate the I67F mutation into rad21 at its endogenous locus and under the control of its endogenous promoter in a strain of *N. crassa* that I will refer to as rad21(I67F). Alongside the mutated copy of the gene I attached a tethered 3xFLAG epitope and the nat-1 gene. I validated that this construct was incorporated and expressed by a PCR assay and western blot respectively. Sanger sequencing also showed that the change responsible for I67F mutation was present in this strain's copy of rad21.

To test whether the mutation resulted in temperature-sensitive lethality I spot-tested conidia at 25˚C and 37˚C. Rad21(I67F) established a faint initial colony at 37 ˚C but this colony stopped growing while the wild type strain was able to continue to grow. After 72 hours, wild type formed a large colony while rad21(I67F) showed no additional growth past its initial size (Figure 7A). To verify that this temperature-sensitive growth defect was not due to the FLAG epitope or something else in the genetic background of rad21(I67F) I performed the same spot-test on rad21-FLAG and strains that are in the background of the mutant. Each of these controls grew at 25˚C and at 37˚C similar to wild type (data not shown).

To further characterize the growth defect of rad21(I67F) I measured its linear growth at 25˚C and after shifting growth conditions to 37˚C. The linear growth rate of the mutant strain was significantly lower than both wild type and rad21-FLAG at 25 ˚C. The decreased fitness of this strain suggests that even at the permissive temperature, cohesin is not fully functional with the RAD21 (I67F) mutation. (Figure 7B). When shifted to 37˚C, rad21 (I67F) stopped growing before running out of space while wild

type and rad21-FLAG continued to grow until they run out of medium and space. (Figure 7C) These data are consistent with the findings of the spot-test, but they indicate that rad21 (I67F) stops growing after about 30 hours at the non-permissive temperature.

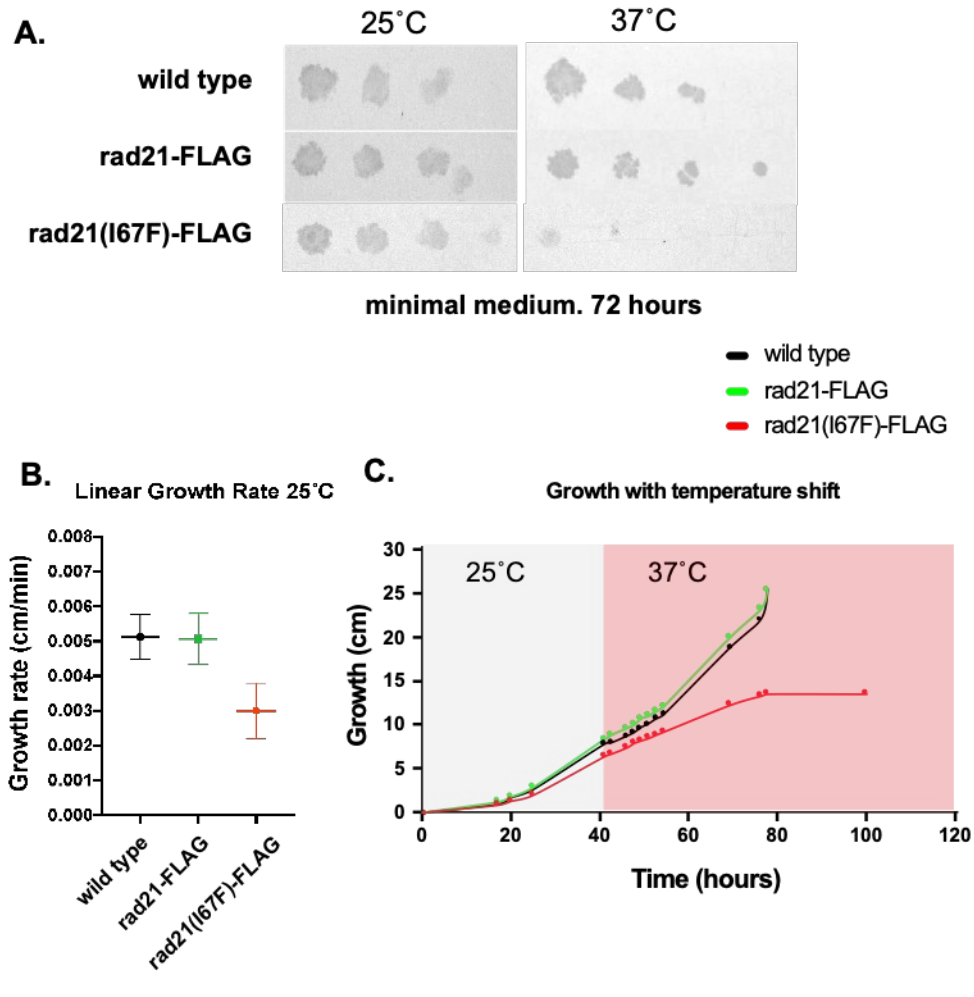


Figure 7: Growth of rad21(I67F) is impaired

A. Spot-tests at 25˚C and at 37˚C on minimal medium after 72 hours. Wild type and rad21-FLAG established at 25˚C and 37˚C while Rad21 (I67F)-FLAG established at 25˚C but was unable to grow beyond a faint initial colonization at 37˚C.
B. Linear growth rates at 25˚C. Linear growth rate of rad21(I67F) was significantly lower than that of wild type and rad21-FLAG strains. ($p=6.2 \times 10^{-7}$)

I additionally sought to establish that normal functions of cohesin are impaired in rad21-(I67F). I performed ChIP on this strain to analyze whether cohesin is lost from cohesin association sites that were identified from psc3-FLAG ChIP-seq data. While rad21(I67F) ChIP-seq data is still in preparation, I tested if sites that were strongly cohesin-enriched (as informed by previous ChIP-seq results) were depleted of cohesin in the rad21(I67F) strain via ChIP qPCR. I designed primers to amplify regions near the transcription termination site of two strongly cohesin marked genes: NCU01638 and NCU04251. I also designed primers to amplify part of a gene, NCU09588, that was very poorly enriched in psc3-FLAG ChIP-seq as a negative control.

In both rad21-FLAG and psc3-FLAG strains, enrichment was strong at NCU01638 and NCU04251 (Figure 8). This is consistent with the literature that members of the cohesin complex are generally enriched at the same genomic loci[54,66] The level of RAD21 was significantly higher in rad21-FLAG than in rad21(I67F)-FLAG ($p = 2.0 \times 10^{-8}$ for NCU01638 and $p = 1.9 \times 10^{-7}$ for NCU04251) and in wild type (Mock FLAG) ($p = 1.1 \times 10^{-8}$ for NCU01638 and $p = 3.3 \times 10^{-7}$ for NCU04251).

These results suggest that cohesin is depleted in rad21(I67F). In addition, they provide further confirmation of the enrichment results obtained from psc3 ChIP-seq and show that components of the cohesin complex localize to the same regions as expected. The high enrichment within psc3-FLAG and rad21-FLAG strains relative to wild type and rad21(I67F)-FLAG strains at the negative control region was surprising. This may indicate that cohesin is enriched at a high basal level across the genome but still localizes to specific association sites where enrichment is higher than this basal level. The high background level within the psc3-FLAG ChIP-seq data and the fact that

cohesin enrichment at NCU01638 and NCU04251 was much higher than at the negative control region support this interpretation. Nevertheless, further analysis will be needed to confirm that cohesin is depleted in the rad21(I67F) strain.

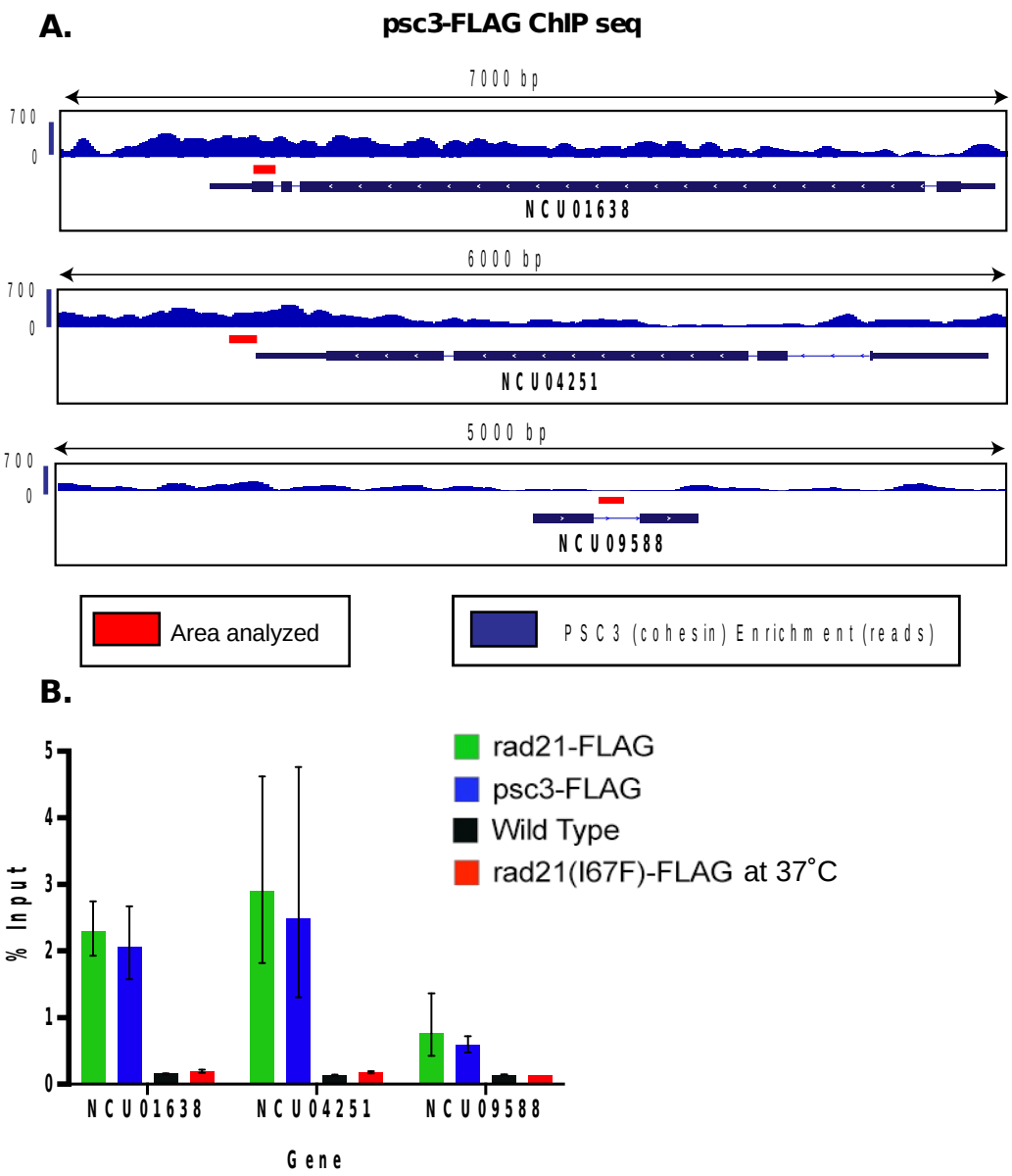


Figure 8: Loss of RAD21(I67F) at cohesin associated sites upon temperature shift

A. Schematic shows enrichment values from ChIP-seq of psc3-FLAG at each of the three loci over which primers were amplified for qPCR. Red boxes underneath enrich - ment tracks indicate the regions that were analyzed via qPCR.
B. ChIP qPCR charts show enrichment (as a percentage of how much of the total DNA was pulled down in the ChIP experiment +/- standard error) for rad21-FLAG, psc3-FLAG, wild type and rad21(I67F)-FLAG strains at NCU01638 and NCU04251 and NCU09588. Rad21(I67F) was germinated at 25˚C and then shifted to 37˚C for 2 hours before ChIP procedure. All other genotypes were cultured at 32˚C.

Discussion

Several important characteristics of the cohesin complex in *N. crassa* were elucidated in this study. Some findings were consistent with the literature in other organisms. In particular, my finding that cohesin associates with the 3' end of transcribed genes and that around half of the intergenic regions with which cohesin associates are between convergent genes, mirrors what was found in *S. pombe*. Here cohesin enrichment partially overlaps with the cohesin loading complex, typically at highly active loci, but, in other cases it is translocated downstream from the cohesin loader due to transcription, similar to what is observed in *S. cerevisae*.[47,48,46] However while nearly all convergent loci are cohesin bound in *S. cerevisae*, 52% of convergent sites are cohesin-bound in *S. pombe* which is more similar to what I have found in *N. crassa*.[48,45] Similarities between the *N. crassa* and *S. pombe* genome that are not shared between *N. crassa* and *S. cerevisiae* such as more frequent introns and larger intergenic sites may influence where cohesin is recruited or where it ultimately localizes.

Strong cohesin binding sites were not observed at centromeres in *N. crassa* via psc3-FLAG ChIP seq. This was unexpected given that cohesin tends to bind to centromeres strongly in *S. cerevisiae, S. pombe*, and *H. sapiens*. The centromere is one of the few places where the mechanism for cohesin recruitment is well established. In *S.pombe*, SWI6 (an HP1 homolog) recruits cohesin while in vertebrates members of the kinetochore complex recruit cohesin in a cell-cycle dependent manner.[59,67] It is possible that the lack of cohesin enrichment in my findings is an artifact of the technique used. In accordance with established procedures, I amplified my ChIP-seq library in order to produce enough material for sequencing. This step is thought to preferentially amplify

GC-rich regions and thus AT-rich regions such as at the centromeres may be under-represented. However ChIP-seq following the same protocol has been used to show enrichment of proteins such as HP1 at centromeres in the past in *N. crassa*.[68] Another possibility is that association with centromeres was not observed because I performed ChIP seq on asynchronous cells. The association with cohesin has been shown to be dynamic in many organisms and these dynamics vary between different organisms. For example in some higher-level eukaryotes the bulk of cohesin is removed from chromosome arms in a manner dependent on Polo-like kinase and WAPL before metaphase, leaving only a fraction at the centromeres that is cleaved during anaphase for chromosome segregation.[15,16] In *S. pombe,* a subset of cohesin remains associated with chromosomes throughout mitosis at both chromosome arms and centromeres and cohesin was shown to spread from heterochromatin to neighboring regions as cells transitioned from metaphase to anaphase.[48] While cohesin strongly associates with centromeres in asynchronous cells and in cells halted before mitosis in both *S. cerevisiae* and *S. pombe,* it is possible that cohesin enrichment at centromeres may be underrepresented in asynchronous cells in *N. crassa.*[46,48]

The successful isolation of a temperature sensitive mutant in cohesin component RAD21 in *N. crassa* came as somewhat of a surprise given that a mutation that is temperature sensitive in one organism (rad21-K1 in *S. pombe*) is not always directly translatable in other organisms. Thus far my findings suggest that cohesin's chromosomal association is impaired or altered in *N. crassa*'s analogous rad21-K1 strain. The addition of more qPCR replicates and ChIP-seq data will enable firmer conclusions. It will be interesting to further characterize this mutant to see if other

phenotypes observed in *S. pombe* such as a defect in mitotic chromosome segregation is also observed.

Overall, I have taken the first steps in characterizing an important protein complex in *N. crassa*. My findings and the strains I developed can serve as a base for future studies into how this protein complex affects various aspects of genomic function in this organism. Specifically, cohesin ChIP-seq data can be cross-referenced to Hi-C data genome wide to determine if cohesin demarcates TADs in *N. crassa*. The rad21(I67F) strain can also be used in Hi-C studies to determine if cohesin plays a role in establishing or maintaining features of genomic architecture such as interacting networks of heterochromatin and TADs. RNA-seq on rad21(I67F) can be used to investigate if cohesin plays a role in gene regulation in *N. crassa* as it does in other organisms. *N. crassa* may serve as a useful and relevant model for studying cohesin and its impact on genomic structure and function in general. Impairing the cohesin complex, a task that involves RNAi or protein based degradation applications in higher organisms is relatively easy with the temperature sensitive-mutant I have isolated.[26,33,69] Additionally, while temperature sensitive mutants also exist in yeast, *N. crassa* is an interesting model to explore because higher organisms share some common features with *N. crassa* that are not shared with yeast.

Materials and Methods

Table 2: N. crassa strains

Strain	Genotype	Source/reference
N1961	Δ*sad-1::hph, mat A*	Shiu, 2001
N2930	Δ*mus-52::bar; his-3; mat A*	Honda and Selker, 2008
N2931	Δ*mus-52::bar; mat a*	Honda and Selker, 2008
N3752	*mat A*	FGSC #2489
N3753	*mat a*	FGSC #4200
N7510	*Eaf3::3xFLAG::nat; Δmus-52::bar; mat A*	Selker lab, unpublished
N7511	*Eaf3::3xFLAG::nat; mat a*	Selker lab, unpublished
N7773	*psc3-3xFLAG::nat; Δmus-52::bar; mat A*	This study
N7774	*psc3-3xFLAG::nat; Δmus-52::bar, mat A*	This study
N8103	*rad21-3xFLAG::nat; Δmus-52::bar, mat a*	This study
N8119	*rad21(I67F)-3xFLAG; mus-52::bar; mat a*	This study
N8120	*rad21(I67F)-3xFLAG; mus-52::bar; mat a*	This study
N8121	*rad21(I67F)-3xFLAG; mus-52::bar; mat a*	This study
N8122	*rad21(I67F)-3xFLAG; mat a*	This study
N8123	*rad21(I67F)-3xFLAG; mus-52::bar; mat a*	This study

Table 3: Oligonucleotides

Primer	Description	Sequence (5' to 3')	Source
6763	*5' FP for psc3::3xFLAG::nat*	AGGCTCGTGTCCTGAGAGACG	This study
6764	*5' RP for psc3::3xFLAG::nat*	*cctccgcctccgcctccgccgcctccgcc*CTCCTCATCCATGTCT TCGTCACC *(nat cassette overlap)*	This study
6765	*3' FP for psc3::3xFLAG::nat*	*gagctcggtaccaagcttgatgcatagc*ATCTGCTTTGGGCGGG CA *(nat cassette overlap)*	This study
6766	*3' RP for psc3::3xFLAG::nat*	CAGCAGAATCAAGGGGAAGAAGAGC	This study
6767	*5' FP for rad21::3xFLAG::nat*	CTTGTGACAGATCCCCTCTTCT	This study
6768	*5' RP for rad21::3xFLAG::nat*	cctccgcctccgcctccgccgcctccgccAGCCGAGGCCTCAATT GGCT *(nat cassette overlap)*	This study
6770	*3' FP for rad21::3xFLAG::nat*	*gagctcggtaccaagcttgatgcatagc*ATCCCGAACTTTGTAG AGTTGGTGTC *(nat cassette overlap)*	This study
6771	*3' RP for rad21::3xFLAG::nat*	TGCTTTATCCATCTGAACTGTACGCTCAT	This study
6772	*5' FP for rad21(I67F)::3xFLAG::nat*	TCCTTGGTGGTTCCTAACATTAGT	This study
6773	*FP for introduction of I67F*	TCTCGGTGTCGTTCGCTTTTATAGTCG	This study

6774	*RP for introduction of I67F*	TTGCGACTATAAAAGCGAACGACACC	This study
4884	*FP for nat cassette*	TCCTTCACCACCGACACCGTCTTCC	Gessaman, 2017
4883	*RP for nat cassette*	AACCCCATCCGCCGGTACGCG	Gessaman, 2017
6757	*NCU09588 qPCR FP*	GGCGTCGGTGAGTTGTGTAA	This study
6758	*NCU09588 qPCR RP*	CTGACGAGAGGGAAAGCGAT	This study
6759	*NCU01638 qPCR FP*	GAAACTACCGTCGGCTTCCT	This study
6760	*NCU01638 qPCR RP*	GTGAAAACGTCGAACGAGCC	This study
6761	*NCU04251 qPCR FP*	AGGACTTGCGTTGTCGTCTT	This study
6762	*NCU04251 qPCR RP*	GGCGGTTCTGTCATGACCTT	This study

Table 4: Other materials

Material Type	**Designation**	**Source/reference**
Antibody (Immunoblot)	*ANTI-FLAG M2-Peroxidase(HRP)*	Sigma (A8592)
Antibody (ChIP)	*ANTI-FLAG M2 Affinity Gel*	Sigma (A2220)

N. crassa culturing.

Liquid cultures (Vogel's minimal medium[70] (VMM), 1.5% sucrose) were inoculated with $10^4 – 10^6$ conidia from *N. crassa* strains at 32 ˚C (or at 25˚C, the permissive temperature or 37˚C, the restrictive temperature, for Rad21 K1 strains) for 1-3 days. Spot-testing was done on bacto-agar plates with VMM, FGS (0.8% sorbose, 0.2% fructose, 0.2% glucose) with or without antiobiotic drug (typically 200 ug/ml Hygromycin B Gold (InvivoGen) or 133ug/ml Nourseothricin (Gold Biotechnology). A hemocytometer was used to count conidia for quantitative spot tests.

Strains were crossed on synthetic crossing medium at 25˚C. Spores were spread on plates, incubated for 1 hour at 60˚C and germinated overnight at 32˚C.

DNA isolation was performed as described previously.[71]

Linear growth rates were assayed as described previously[72] with the exception that disposable plastic tubes were used in lieu of glass ones. One set of tubes was kept at 25˚C throughout analysis while another set was grown at 25˚C for ~2 days and then shifted to 37˚C.

Tagged strain construction

C-terminal 10xGly-3xFLAG constructs were amplified and inserted into my genes of interest in line with previous work[42]. Primers were designed to amplify two 1000 base pair gene segments from genomic DNA, one within the open reading frame of the gene and the other 500 base pairs downstream of the gene's stop codon. Primers included overlap with nat-1 cassette (10xGly-3xFLAG and nat-1 conferring resistance to the antibiotic nourseothricin from plasmid FJ457009.1) and were incorporated at the C-terminal end of the gene during transformation via homologous recombination. Constructs were transformed into competent strains via electroporation to their endogenous locus. Incorporation of the construct into heterokaryon strains was validated by Southern blot or PCR amplification and protein expression in primary transformants was verified by western blot. Homokaryons were generated by back-crossing crossing to wildtype (N3753 or N3752).

Mutant strains construction

An approach, similar to the method described above was taken for construction of rad21(I67F) strains except that primers 6773 and 6774 were designed to introduce a point mutation in the gene body to change the 67th amino acid from an isoleucine to a phenylalanine thereby recapitulating the Rad21K1 mutation. In addition to PCR assay

and Western blotting, Sanger sequencing verified that point mutation responsible for I67F change was incorporated. Strains were crossed to N1961 (Δsad2) to generate homokaryons.

Southern blot

600 ng of genomic DNA were used and digested by appropriate restriction digest enzyme for 1 hour at 37 ˚C. DNA was fractionated and separated via gel-electrophoresis, transferred to nitrocellulose membrane and crosslinked. Membranes were hybridized to probes containing randomly incorporated ^{32}P. For genotyping, NAT-1 or Hygromycin resistance gene probes were used. Membrane radiation was transferred to a GE phosphoscreen and imaged on a STORM 860 Phospohorimager (Molecular Dynamics).

Western blot

Western blots were performed as previously described.[71] Conidia were cultured in VMM and 1.5% sucrose overnight. Tissue was collected 500 µl ice cold lysate buffer (50 mM HEPES [pH 7.5], 150 mM NaCl,10% glycerol, 0.02-0.2% NP-40, and 1 mM EDTA supplemented with 1x Halt TM Protease Inhibitor Cocktail (Thermo Scientific)). ANTI-FLAG M2 Peroxidase (HRP) was used as an antibody (Sigma A8592).

Chromatin immunoprecipitation (ChIP) preparation and ChIP qPCR

ChIP preparation was prepared as described previously[73] with some modifications. Approximately 1.0 x 10^6 conidia were inoculated in 5 ml VMM with 1.5% sucrose at 32˚C with shaking overnight (about 16 hours). For analysis of rad21(I67F) strains cultures were grown at 25˚C instead of 32˚C and shifted to 37˚C for

2 hours in accordance with analysis of rad21K1 in *S. pombe*[4]. Crosslinking was done with 37% formaldehyde for 30 minutes with shaking at room temperature. Chromatin was sheared for a total of 20 minutes with Diagenode Bioruptor at high power. Samples inclubated with ANTI-FLAG M2 Affinity Gel (Sigma A2220) overnight. ChIP DNA for qPCR was cleaned up with PCR cleanup kit using MinElute® columns (Qiagen) and eluted in Qiagen elution buffer. qPCR was performed with Quanta Biosciences PerfeCTa® Sybr Green FastMix on an Applied Biosystems Step One Plus Real-Time PCR platform. Inputs were diluted 1:20 and primers used are listed in the Oligos table. Results were visualized using GraphPad Prism.

Sequencing library prep and analysis

Library preparation was done in accordance with previously described procedures.[74,75] Approximately 10 ng of ChIP DNA was used to generate libraries using Illumina TruSeq kits A and B (Illumina IP-202-1012 and IP202-1024). Libraries were amplified for 1 cycle of 30 second at 98˚C, 18 cycles of 10 seconds at 98˚C followed by 30 seconds at 60˚C and 30 seconds at 72˚C and then 1 cycle of 5 minutes at 72˚C. Libraries were sequenced on an Illumina Next Seq 500 platform which generates 75 base pair single-end reads. Sequencing analysis was performed using the DeepTools2 suite from open-source platform Galaxy.[63,76] Reads were mapped to the corrected *Neurospora crassa* OR74A genome (NC12) using Bowtie2.[43,77] BamCoverage was used to generate Bigwig files from aligned reads in BAM format. BamCompare was used to generate $\log_2$ratios between BAM files for normalization of data. ComputeMatrix was used to determine reads over predetermined sequence files in BED format. PlotProfile and PlotHeatmap were used to visualize matrix data.

BED files for gene pairs in convergent, divergent and same direction orientations were generated using a series of python scripts. First each gene in *N. crassa*'s genome was categorized by its orientation in reference to the gene upstream to it in grab_other_pair.py. Convergentpairs.py was used to cross-reference information about the orientation of each gene from the previous script with a list of every gene pair (generated in Microsoft Excel) to create a list of all convergent gene pairs. Analogous scripts were used for each of the other three possible gene orientations.

Bibliography

1. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).

2. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

3. Belton, J.-M., McCord, R. P., Gibcus, J. H., Naumova, N. & Zhan, Y. Hi–C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).

4. Mizuguchi, T. *et al.* Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature* **516**, 432–435 (2014).

5. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

6. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

7. Rowley, M. J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* **67**, 837-852. (2017).

8. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).

9. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).

10. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

11. Lajoie, B. R., Dekker, J. & Kaplan, N. The Hitchhiker's guide to Hi-C analysis: Practical guidelines. *Methods* **72**, 65–75 (2015).

12. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).

13. Smith, J. S., Savage, K. I., Thompson, A. & Mills, K. I. Loss of Function Cohesin Complex Gene Mutations Create Neomorphic Cell States Advantageous to Oncogenesis. *Blood* **128**, (2016).

14. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).

15. Losada, A., Hirano, M. & Hirano, T. Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes Dev.* **12**, 1986–1997 (1998).

16. Darwiche, N., Freeman, L. A. & Strunnikov, A. Characterization of the components of the putative mammalian sister chromatid cohesion complex. *Gene* **233**, 39–47 (1999).

17. Peters, J. M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes and Development* **22**, 3089–3114 (2008).

18. Haering, C. H., Löwe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* **9**, 773–88 (2002).

19. Deardorff, M. A., Noon, S. E. & Krantz, I. D. *Cornelia de Lange Syndrome*. *GeneReviews*® (University of Washington, Seattle, 1993).

20. Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J. & Strachan, T. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat. Genet.* **36**, 636–641 (2004).

21. Krantz, I. D. *et al.* Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. *Nat. Genet.* **36**, 631–635 (2004).

22. Dorsett, D. & Merkenschlager, M. Cohesin at active genes: a unifying theme for cohesin and gene expression from model organisms to humans. *Curr. Opin. Cell Biol.* **25**, 327–333 (2013).

23. Remeseiro, S. *et al.* Reduction of Nipbl impairs cohesin loading locally and affects transcription but not cohesion-dependent functions in a mouse model of Cornelia de Lange Syndrome. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1832**, 2097–2102 (2013).

24. Kaur, M. *et al.* Precocious sister chromatid separation (PSCS) in Cornelia de Lange syndrome. *Am. J. Med. Genet.* **138 A**, 27–31 (2005).

25. Heidinger-Pauli, J. M., Mert, O., Davenport, C., Guacci, V. & Koshland, D. Systematic Reduction of Cohesin Differentially Affects Chromosome Segregation, Condensation, and DNA Repair. *Curr. Biol.* **20**, 957–963 (2010).

26. Pauli, A. *et al.* Cell-Type-Specific TEV Protease Cleavage Reveals Cohesin Functions in Drosophila Neurons. *Dev. Cell* **14**, 239–251 (2008).
27. Schuldiner, O. *et al.* piggyBac-Based Mosaic Screen Identifies a Postmitotic Function for Cohesin in Regulating Developmental Axon Pruning. *Dev. Cell* **14**, 227–238 (2008).

28. Pauli, A. *et al.* A direct role for cohesin in gene regulation and ecdysone response in drosophila salivary glands. *Curr. Biol.* **20**, 1787–1798 (2010).

29. Lee, B.-K. & Iyer, V. R. Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. *J. Biol. Chem.* **287**, 30906–13 (2012).

30. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).

31. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32**, 3119–29 (2013).

32. Kim, K.-D., Tanizawa, H., Iwasaki, O. & Noma, K. Transcription factors mediate condensin recruitment and global chromosomal organization in fission yeast. *Nat. Genet.* **48**, 1242–1252 (2016).

33. Rao, S. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171(2)**, 305-320 (2017).

34. Nativio, R. *et al.* Cohesin Is Required for Higher-Order Chromatin Conformation at the Imprinted IGF2-H19 Locus. *PLoS Genet.* **5**, e1000739 (2009).

35. Chien, R. *et al.* Cohesin Mediates Chromatin Interactions That Regulate Mammalian β-globin Expression. *J. Biol. Chem.* **286**, 17870–17878 (2011).

36. Seitan, V. C. *et al.* A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* **476**, 467–471 (2011).

37. Srinivasan, M. *et al.* The Cohesin Ring Uses Its Hinge to Organize DNA Using Non-topological as well as Topological Mechanisms. *Cell* **173(6)**, 1508-1519 (2018).
38. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.* **112**, 6456–6465 (2015).

39. Aramayo, R. & Selker, E. U. Neurospora crassa, a model system for epigenetics research. *Cold Spring Harb. Perspect. Biol.* **5**, 1–18 (2013).

40. Perkins, D. D. & Turner, B. C. Neurospora from natural populations: Toward the population biology of a haploid eukaryote. *Experimental Mycology* **12**, 91–131 (1988).

41. Colot, H. V *et al.* A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10352–10357 (2006).

42. Honda, S. & Selker, E. U. Tools for fungal proteomics: multifunctional neurospora vectors for gene replacement, protein expression and protein purification. *Genetics* **182**, 11–23 (2009).

43. Galazka, J. M. *et al.* Neurospora chromosomes are organized by blocks of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. *Genome Res.* **26**, 1069–80 (2016).

44. Klocko, A. D. *et al.* Normal chromosome conformation depends on subtelomeric facultative heterochromatin in Neurospora crassa. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 15048–15053 (2016).

45. Glynn, E. F. *et al.* Genome-Wide Mapping of the Cohesin Complex in the Yeast Saccharomyces cerevisiae. *PLoS Biol.* **2**, 259 (2004).

46. Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**, 573–8 (2004).

47. Gullerova, M. & Proudfoot, N. J. Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. *Cell* **132**, 983–95 (2008).

48. Schmidt, C. K., Brookes, N. & Uhlmann, F. Conserved features of cohesin binding along fission yeast chromosomes. *Genome Biol.* **10**, R52 (2009).

49. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).

50. Rubio, E. D. *et al.* CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8309–8314 (2008).

51. Parelho, V. *et al.* Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell* **132**, 422–433 (2008).

52. Misulovin, Z. *et al.* Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. *Chromosoma* **117**, 89–102 (2008).

53. Tatebayashi, K., Kato, J. & Ikeda, H. Isolation of a Schizosaccharomyces pombe rad21ts Mutant That Is Aberrant in Chromosome Segregation, Microtubule Function, DNA Repair and Sensitive to Hydroxyurea: Possible Involvement of Rad21 in Ubiquitin-Mediated Proteolysis. *Genetics* **148**, (1998).

54. Tomonaga, T. *et al.* Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. *Genes Dev.* **14**, 2757–70 (2000).

55. Toyoda, Y. *et al.* Requirement of Chromatid Cohesion Proteins Rad21/Scc1 and Mis4/Scc2 for Normal Spindle-Kinetochore Interaction in Fission Yeast either by microtubule-associating proteins or microtubule dynamic instability. *Current Biology* **12**, (2002).

56. Adachi, Y., Kokubu, A., Ebe, M., Nagao, K. & Yanagida, M. Cut1/separase-dependent roles of multiple phosphorylation of fission yeast cohesin subunit Rad21 in post-replicative damage repair and mitosis. *Cell Cycle* **7**, 765–776 (2008).

57. Sonoda, E. *et al.* Scc1/Rad21/Mcd1 Is Required for Sister Chromatid Cohesion and Kinetochore Function in Vertebrate Cells. *Dev. Cell* **1**, 759–770 (2001).

58. Kong, X. *et al.* Cohesin associates with spindle poles in a mitosis-specific manner and functions in spindle assembly in vertebrate cells. *Mol. Biol. Cell* **20**, 1289–301 (2009).

59. Hinshaw, S. M., Makrantoni, V., Harrison, S. C. & Marston, A. L. The Kinetochore Receptor for the Cohesin Loading Complex. *Cell* **171**, 72-84.e13 (2017).

60. Xu, X. *et al.* Suppressor mutation analysis combined with 3D modeling explains cohesin's capacity to hold and release DNA. *Proc Natl. Acad. Sci.* **115(21)**, 4833-4842 (2018).

61. Hopp, T. P. *et al.* A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio/Technology* **6**, 1204–1210 (1988).

62. Dorsett, D. & Misulovin, Z. Measuring Sister Chromatid Cohesion Protein Genome Occupancy in Drosophila melanogaster by ChIP-seq. *Methods Mol. Biol.* **1515**, 125–139 (2017).

63. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, 537–544 (2018).

64. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488 (2008).

65. Pareek, M. *et al.* Alternative Functional rad21 Paralogs in Fusarium oxysporum. *Front. Microbiol.* **10**, (2019).
66. Misulovin, Z. *et al.* Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. *Chromosoma* **117**, 89–102 (2008).

67. Nonaka, N. *et al.* Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nat. Cell Biol.* **4**, 89–93 (2002).

68. Honda, S. *et al.* Heterochromatin protein 1 forms distinct complexes to direct histone deacetylation and DNA methylation. *Nat. Struct. Mol. Biol.* **19**, 471–7, S1 (2012).

69. Vass, S. *et al.* Depletion of Drad21/Scc1 in Drosophila Cells Leads to Instability of the Cohesin Complex and Disruption of Mitotic Progression. *Curr. Biol.* **13**, 208–218 (2003).

70. Davis, R. H. & de Serres, F. J. [4] Genetic and microbiological research techniques for Neurospora crassa. *Methods Enzymol.* **17**, 79–143 (1970).

71. Honda, S. & Selker, E. U. Direct interaction between DNA methyltransferase DIM-2 and HP1 is required for DNA methylation in Neurospora crassa. *Mol. Cell. Biol.* **28**, 6044–55 (2008).

72. Ryan, F. J., Beadle, G. W. & Tatum, E. L. THE TUBE METHOD OF MEASURING THE GROWTH RATE OF NEUROSPORA. *Am. J. Bot.* **30**, 784–799 (1943).

73. Jamieson, K. *et al.* Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. *Genome Res.* **26**, 97–107 (2016).

74. Bicocca, V. T., Ormsby, T., Adhvaryu, K. K., Honda, S. & Selker, E. U. ASH1-catalyzed H3K36 methylation drives gene repression and marks H3K27me2/3-competent chromatin. *Elife* **7**, (2018).

75. Jamieson, K. *et al.* Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. *Genome Res.* **26**, 97–107 (2016).

76. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, 160–165 (2016).

77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.*

Methods **9**, 357–359 (2012).